Ontology for Earth System Science: Best Practices

Rob Raskin
JPL

Data to Knowledge

	Data		Information		Knowledge	
Basic Elements	Bytes	Numbers	Models		Facts	
Storage	File	Database	GIS		Ontology	Mind
Volume	High				Low	
Density	Low				High	
Services	Save	Discover	Visualize	Overlay	Infer	Understand Predict

Syntax → ***Semantics***

What is Knowledge?

- Facts, relations, meanings
- Information with context
  - *Information is summarized or organized data*
- Common sense
- Shared understanding
- Suitable for reasoning/inference
- Dynamic, expandable
- Requires semantics

Semantic Understanding is Difficult (even for humans)!

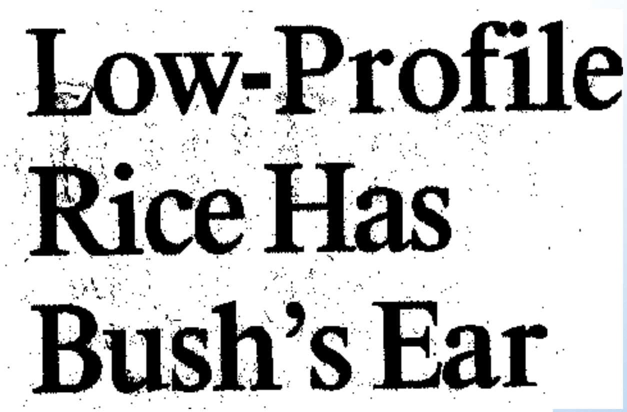
Variable t: temperature — Data quality= 5

Variable t: time — Data quality= 3

"not a nobody" *(double negative=positive)*
"ain't nobody" *(single negative empahsized)*
"yeah, right" *(double positive= negative)*

Let's eat, Grandma.
Let's eat Grandma.

Time flies like an arrow.
Fruit flies like a pie.


Low-Profile
Rice Has
Bush's Ear

LA Times headline

"Mission accomplished. Major combat operations in Iraq have ended"

-Pres. Bush, 2003

What is an ontology?

- A dictionary
  - in a form that is understandable both by computers and humans
- A namespace
  - A URL/URI that contains an authoritative declaration of a concept
- Shared understanding of concepts
  - A reference of knowledge

Why point to an ontology?

- To remove ambiguity. Examples:
  - Ocean surface wind
    - measured at what height?
  - Temperature anomaly:
    - relative to what climatological average?
  - Measurement under clear skies
    - based upon what definition of clear?
- To capture provenance. Example:
  - Using moisture adjustment model
    - which model?

Ontology Characteristics

- Concept space
- Dynamic
  - Knowledge is not static
- Use of a standard language makes it easy to extend (specialize) concepts developed by others
- Synonym support (multiple terms with same meaning)
  - Label available to indicate preferred term for each community
- Homonym support (multiple meanings of same term)
  - Separate namespaces (President:Bush vs Plant:Bush)

Controlled Vocabulary Characteristics

- GCMD science keywords
  - Hierarchical subject classification
  - Focused on collection descriptions rather than individual parameter listings
  - Approx 1200 names
- CF parameter names
  - Enables qualified extensions
    - Air_temperature_under_clear_skies
  - Requires moderator approval for additions
    - Discussion list provides feedback before acceptance
  - Approx. 800 names, increasing rapidly

Compared to a Controlled Vocabulary, an Ontology is…

- More scalable
  - Easy to mix and match combination of terms
- More descriptive
  - Full definitions in terms of other terms
- Multi-faceted
  - Combination of parameters possible

Use Case: Global Warming Query

Find data which demonstrates global warming at high latitudes during summertime and plot warming rate.

- Extract information from the use-case – encode knowledge
- Translate this into a complete query for data – inference and integration of data from instruments, indices and models

"Global warming"= Trend of increasing temperature over large spatial scales

"High latitude"= Latitudes > 60 degrees

"Summertime"= June-Aug (NH) and Jan-Mar (SH)

"Find data"= Locate datasets using catalogs, then access and read it

"Plot warming rate"= Display temperature vs time

Ontology Representations…

- As XML
- As Triples
- In DBMS
- Visually

Ontology: Triple Representation

Subject–Verb–Object representation

Flood	is a	WeatherPhenomena
GeoTIFF	is a	FileFormat
Soil Type	is a	PhysicalProperty
Pacific Ocean	is a	Ocean

Ocean	has substance	Water
Sensor	measures	Temperature

Ontology: Visual Representation

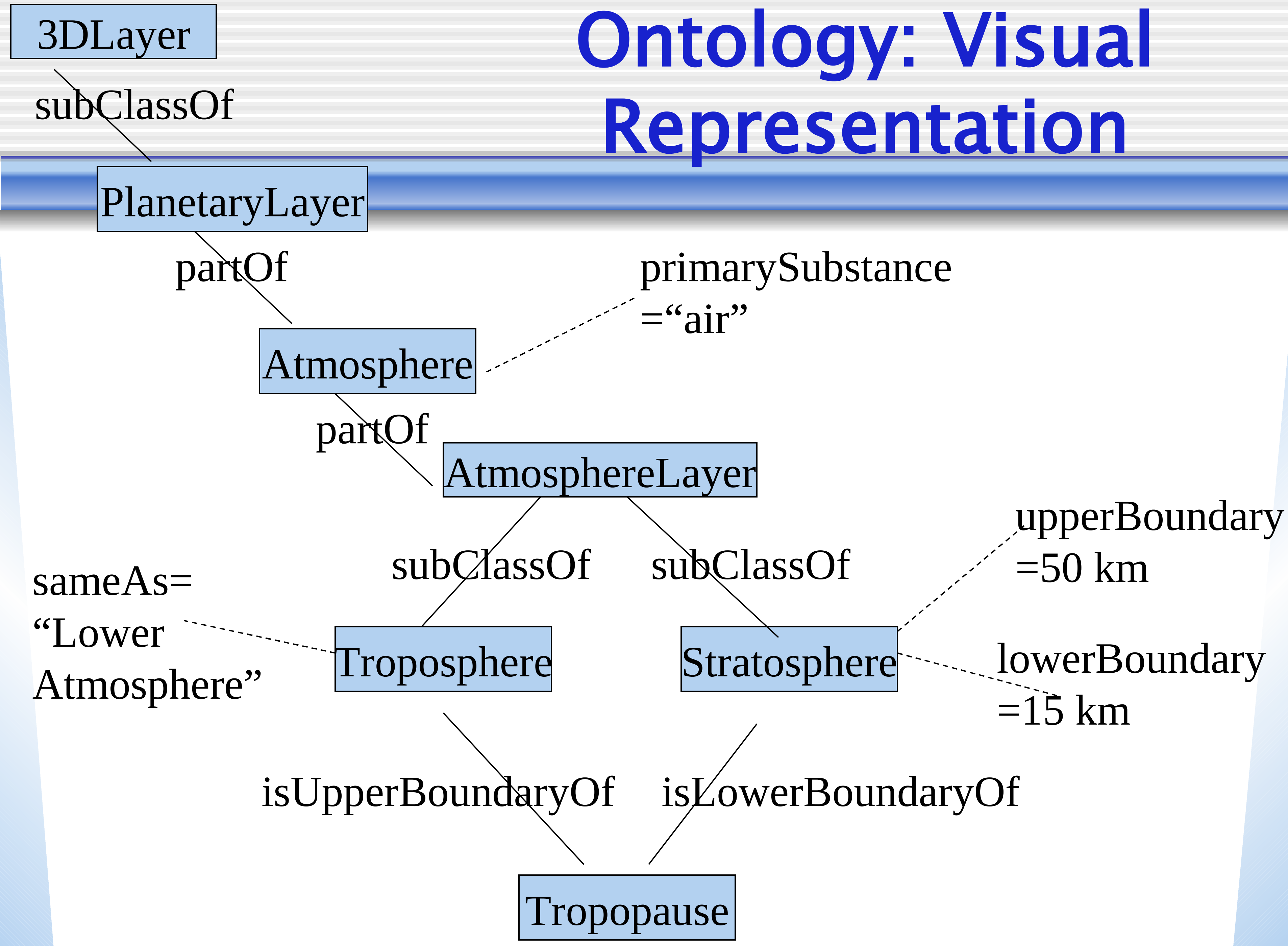
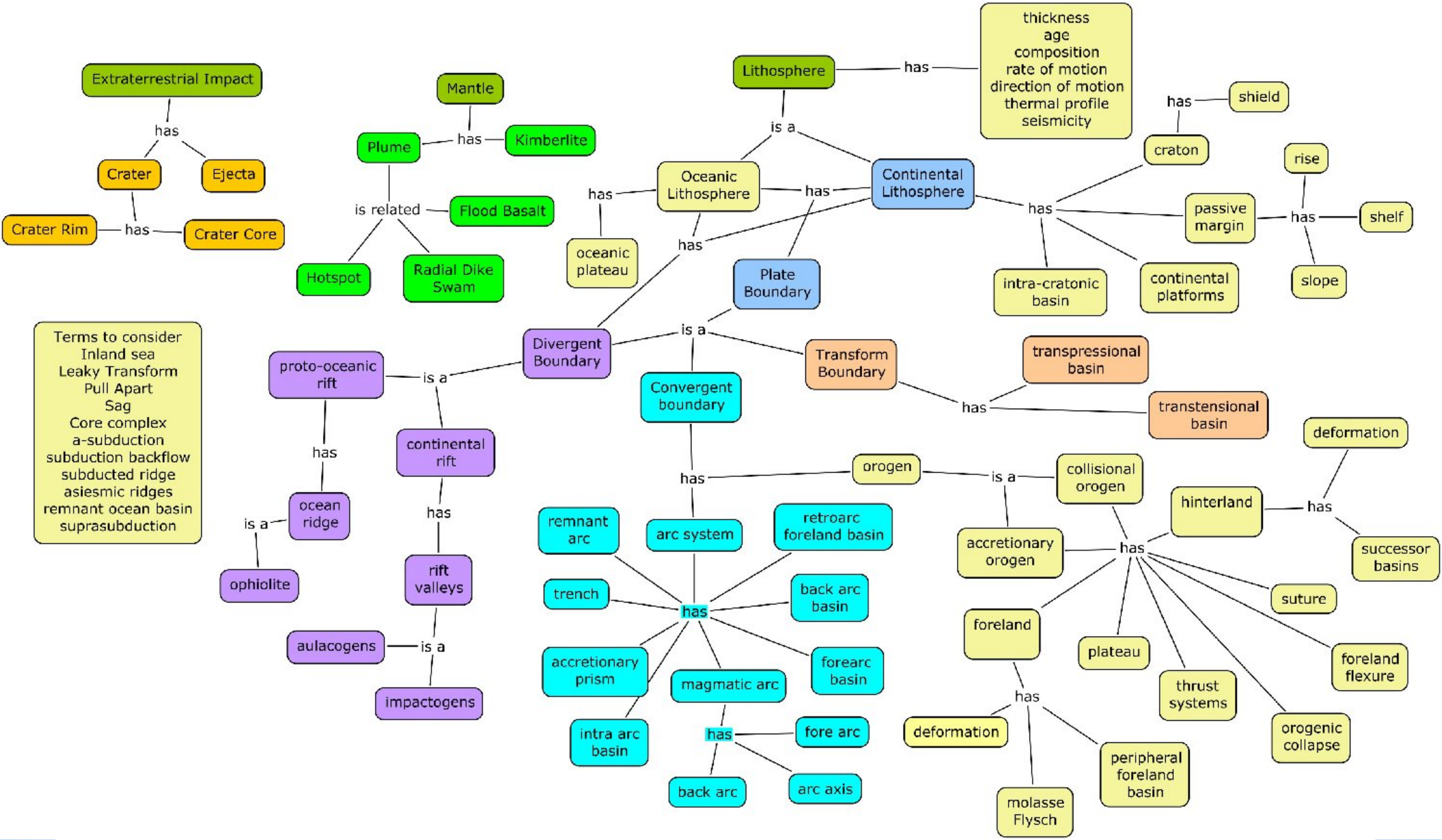
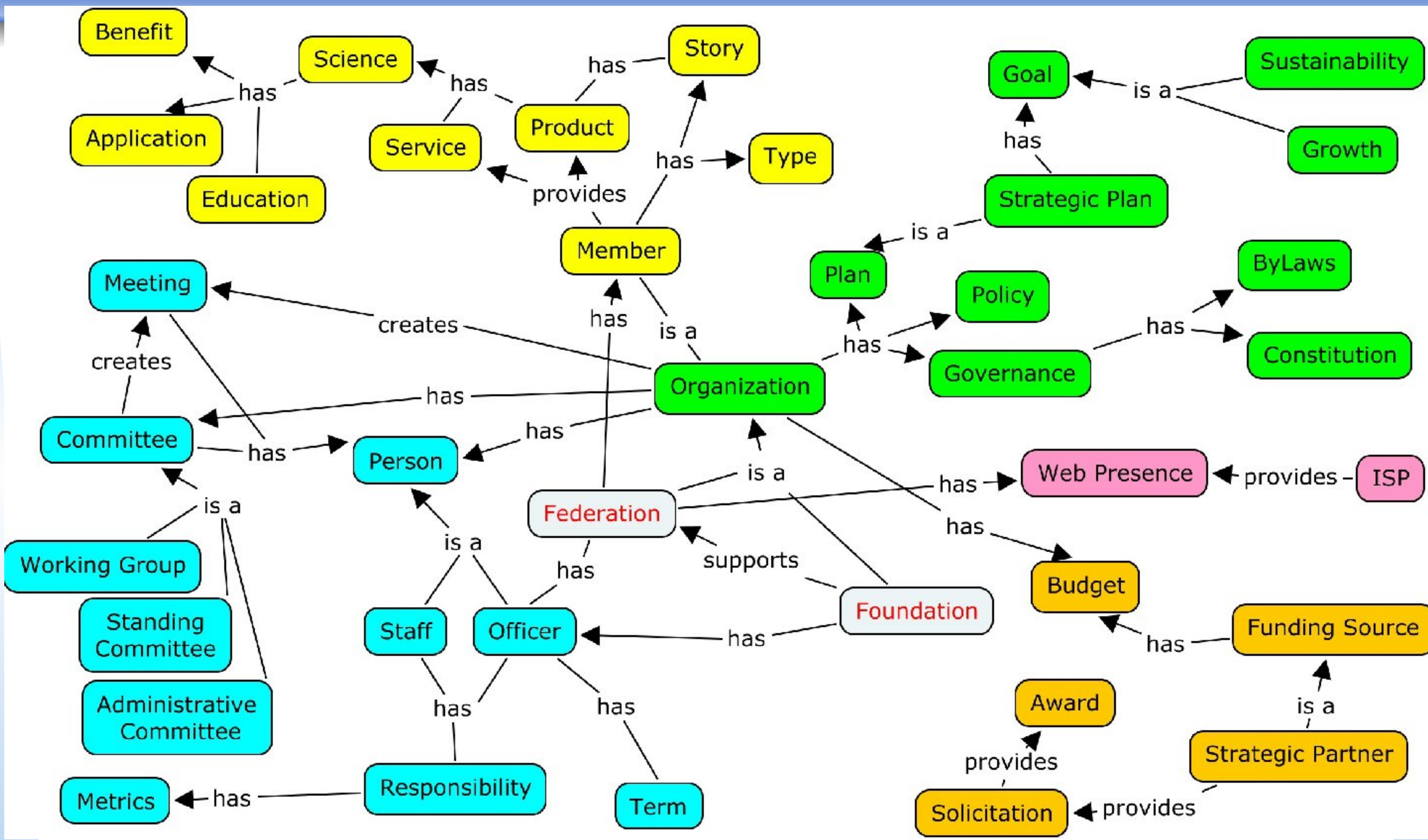


Plate Tectonics Ontology



Ontology of ESIP Federation



Ontology: XML Representation

```xml
<owl:Class rdf:ID="StormSurge">
  <rdfs:subClassOf rdf:resource="#Flood"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasAssociatedEarthRealm"/>
      <owl:allValuesFrom rdf:resource="#earthrealm.owl#CoastalRegion"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>

<owl:Class rdf:ID="Flood">
  <rdfs:subClassOf rdf:resource="#SevereWeatherPhenomena"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasAssociatedEarthRealm"/>
      <owl:allValuesFrom rdf:resource="earthrealm.owl#LandSurface"/>
    </owl:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#hasAssociatedSubstance"/>
      <owl:allValuesFrom rdf:resource="substance.owl#LiquidWater"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

XML-based Ontology Languages

- XML satisfies desired properties for language *syntax*
  - Readable by both humans and machines
- However, there are too many possible ways that XML tags can be named and used
- No standardization of XML tag meanings as in HTML (pair => renders in bold)
- Additional standardized semantics needed to exploit shared understanding of concepts

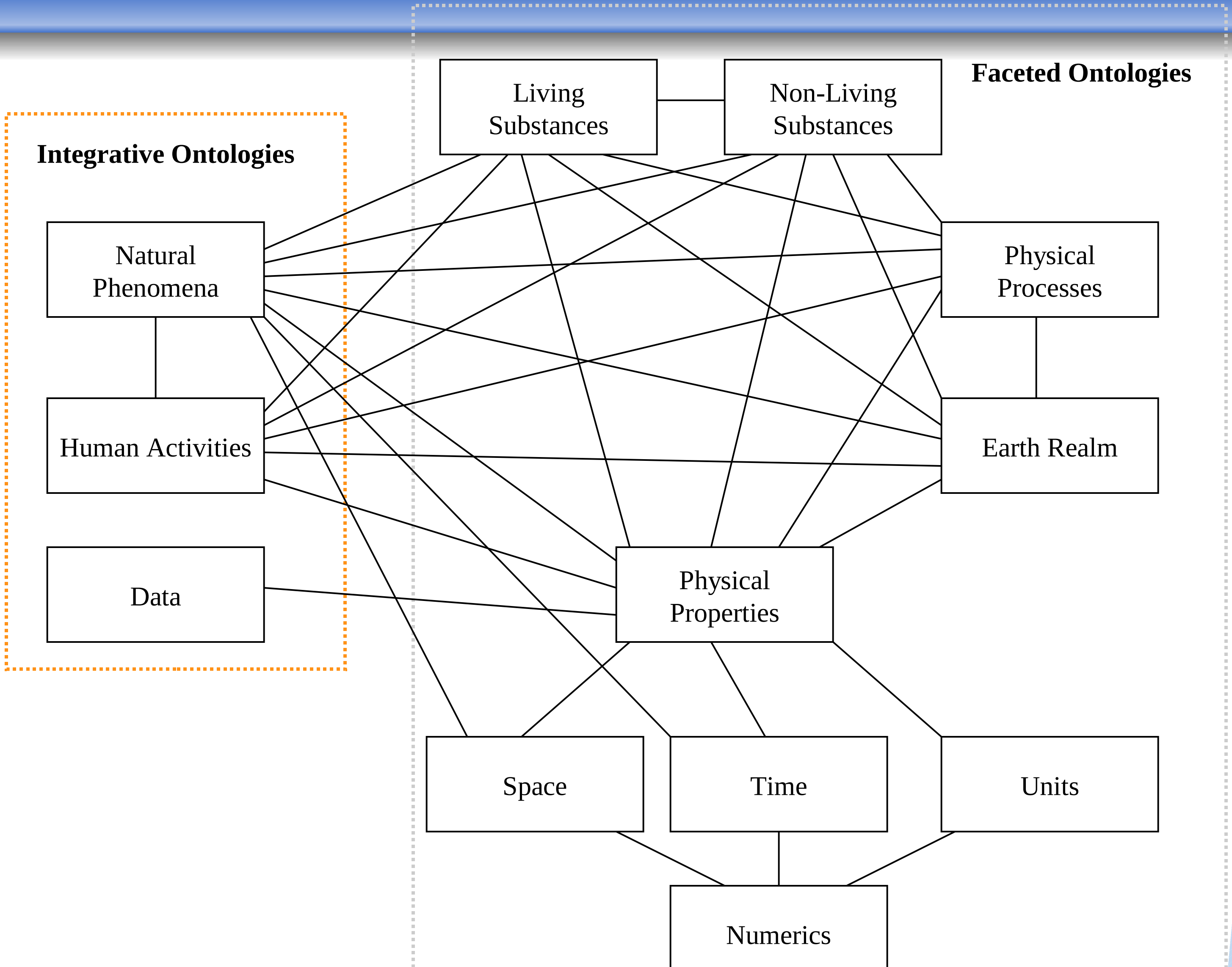
Ontology Languages in XML: RDF and OWL

- Use of standard languages make it easy to extend (specialize) concepts developed by others
- World Wide Consortium (W3C) has adopted languages that specialize XML
  - Resource Description Formulation (RDF)
    - Standardizes: class, subclass, property, subproperty
  - Ontology Web Language (OWL)
    - Standardizes: transitive, inverse relations, cardinality, etc.

SWEET

- Semantic Web for Earth and Environmental Terminology (SWEET)
- Includes concepts of:
  - Earth system science
  - Data

SWEET 1.0 Ontologies



SWEET Faceted Science Ontologies

- Earth Realms
  - Atmosphere, SolidEarth, Ocean, LandSurface, …
- Physical Properties
  - temperature, composition, area, albedo, …
- Substances
  - CO2, water, lava, salt, hydrogen, pollutants, …
- Living Substances
  - Humans, fish, …

SWEET Integrative Science Ontologies

- Phenomena
  - ElNino, Volcano, Thunderstorm, Deforestation, Terrorism, physical processes (e.g., convection)
  - Each has associated EarthRealms, PhysicalProperties, spatial/temporal extent, etc.
  - Specific instances included
    - e.g., 1997-98 ElNino
- Human Activities
  - Sustainability, Fisheries, IndustrialProcessing, Economics, …

SWEET Numerical Ontologies

- Intervals, numeric relations (<, >)
- Cartesian products
- Functions, derivatives
- Statistical concepts
  - probability density functions
- Fuzzy concepts
  - "near"
- Spatial concepts
  - 0-D, 1-D, 2-D, and 3-D objects
  - Coordinate systems
  - Above, inside, etc.
- Temporal concepts
  - Instant, durations, geological time scales
- Decision/risk concepts

SWEET Data Ontology

- Dataset characteristics
  - Format, data model, dimensions, …
- Provenance
  - Source, processing history, …
- Parameters
  - Scale factors, offsets, …
- Data services that make sense
  - Subsetting, reprojection, …
- Quality measures
- Special values
  - Missing, land, sea, ice, ...

SWEET 2.0 Modular Design

- Supports easy extension by domain specialists
- Organized by subject (theoretical to applied)
- Reorganization of classes, but no significant changes to content
- Importation is unidirectional

Math, Time, Space

Basic Science

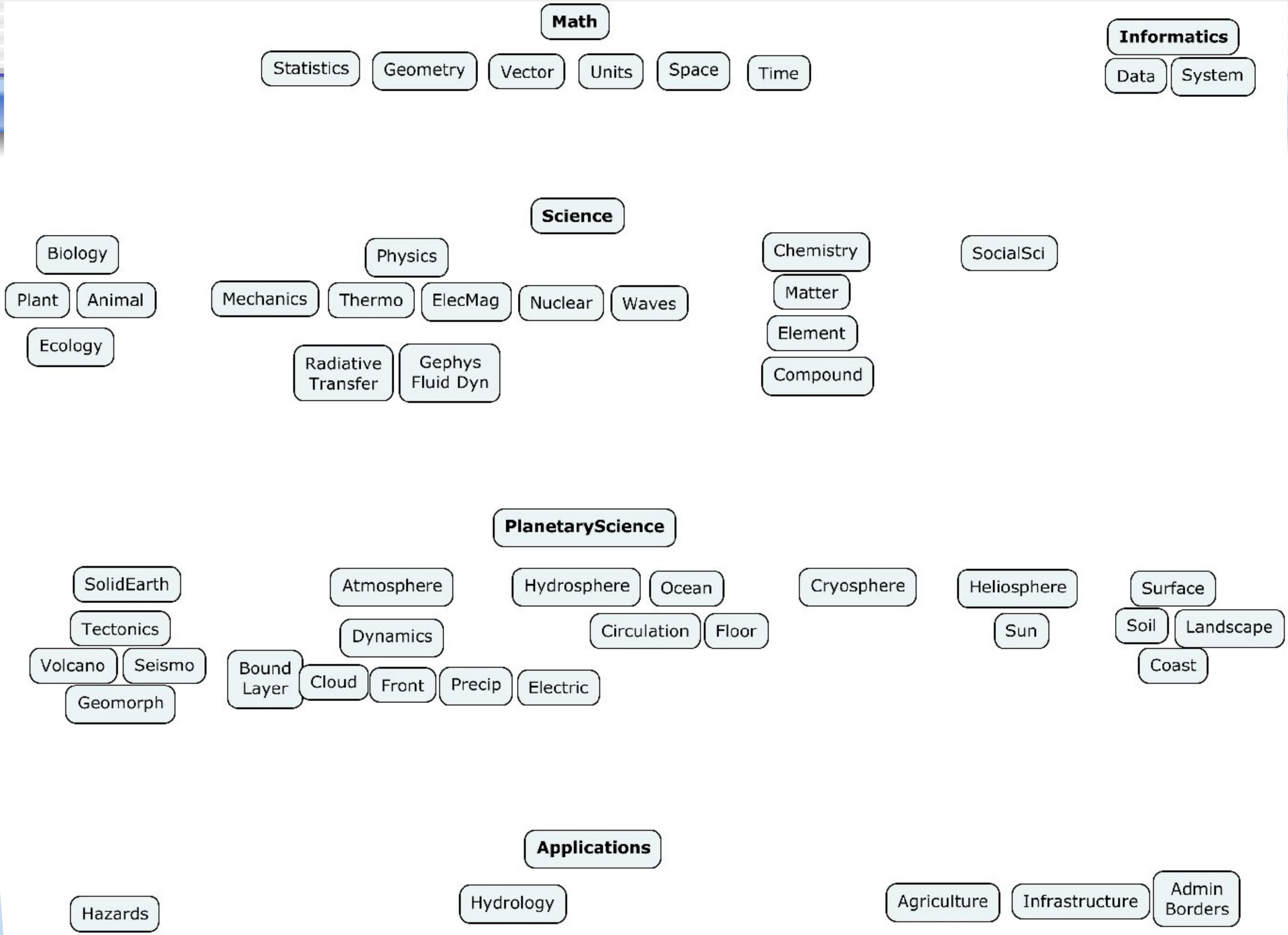
Geoscience Processes

Geophysical Phenomena

Applications

importation

SWEET 2.0 Ontologies



SWEET 2.0 New Features

- Organized by subject
  - Makes it easy for domain specialists to add new modules
- Smaller, modular ontologies
- 12 ontologies -> 80 ontologies

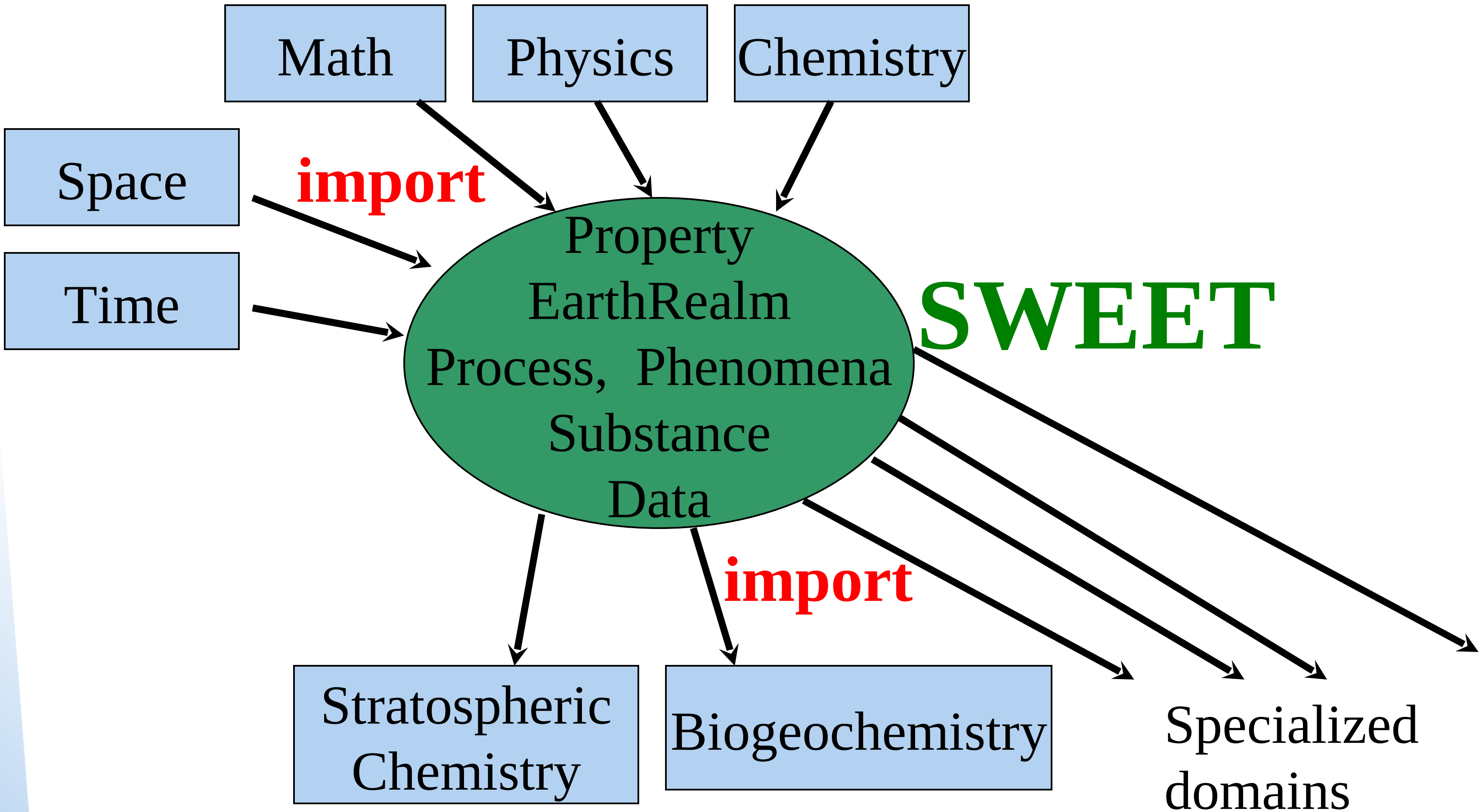
Best Practices

- Keep ontologies small, modular
  - Be careful that “Owl:Import” imports everything
  - Use higher level ontologies where possible
- Identify hierarchy of concept spaces
  - Model schemas
  - Try to keep dependencies unidirectional

Consistent OWL Representations

- IntervalQuantity
  - hasLowerBound
  - hasUpperBound
  - hasUnit
- Fuzzy concepts
  - "nearlySameAs"
  - relatedConcept connects siblings or other concepts that might be closely related
  - similarity matrix provides more precise support for search results ranking
- Uncertainty
  - pdf representations

SWEET as an Upper Level Earth Science Ontology



Why an Upper-Level Ontology for Earth System Science?

- Many common concepts used across Earth Science disciplines (such as properties of the Earth)
  - Provides common definitions for terms used in multiple disciplines or communities
  - Provides common language in support of community and multidisciplinary activities
  - Provides common “properties” (relations) for tool developers
- Reduced burden (and barrier to entry) on creators of specialized domain ontologies
  - Only need to create ontologies for incremental knowledge

Collaborative Ontology Development

Planetary Ontologies

Use Cases – Federation ... | Planetary Ontologies

Ontology

Planetary Ontologies

A web site for sharing ontologies in the Earth and Space Sciences

Purpose and Scope

- To facilitate the use, re-use, development, evolution, alignment and merging of ontologies by a community of best-practices in the broad fields of Earth and Space Sciences.
- To enable collaborative use and discussion of ontologies by practitioners.

Ontology Collaboration

- Ontologies
- RSS Feed for ontology changes
- Wiki
- Mailing List

Links

- SESDI (Semantically-Enabled Science Data Integration)
- SWEET (Semantic Web for Earth and Environmental Terminology)
- GEON (Geosciences Network)
- VSTO (Virtual Solar-Terrestrial Observatory)
- NOESIS (Earth and Atmospheric Science Smart Search)
- MMI (Marine Metadata Initiative)
    - VINE (Vocabulary Integration Environment Tool)

Private

- SWEET mailing list

Planetont.org is a community ontology project and forum developed with support from NASA.

Ontology Development Tools: CMAP

- Free, downloadable tool for knowledge representation and ontology development
- Visual language with input/export to OWL
  - Supports subset of OWL language
- http://cmap.ihmc.us/coe

Resources

- SWEET
  - http://sweet.jpl.nasa.gov
- Ontology development/sharing site
  - http://PlanetOnt.org